Queen's University
Faculty of Arts and Sciences
Department of Economics
Fall 2009

Economics 250 A: Introduction to Statistics

Midterm Exam I:
Time allotted: 80 minutes.

Instructions:

READ CAREFULLY. Calculators are permitted. At the end of the exam are several formulae, the cumulative binomial and standard normal distribution. Answers are to be written in the examination booklet. Do not hand in the question sheet. You are to answer **ALL** questions. **SHOW ALL YOUR WORK**. Most of the marks are awarded for showing how a calculation is done and not for the actual calculation itself! There are a total of 100 possible marks. Answer all ten (10) questions (marks are indicated)

Part I (50 marks) There are 5 questions in total for this part. This part examines discrete and continuous probabilities

It has been determined that a full 2 second stop is required legally at all stop signs for safety considerations. Motorists are equally like to stop between 0 and 10 seconds at any stop sign. A traffic cop is secretly watching in the bushes and has a quota of 4 tickets for too short a stop. Answer the following questions using this information making any assumptions you require.

1. **(10 marks)** What is the probability of any person arriving at the stop sign making an illegal stop? Show the probability in a diagram and explain the distribution you are using.

$$\begin{aligned} S &\sim \text{Uniformly} \\ P(0 &< S < 2) = \frac{1}{10} \times (2-0) = .2 \end{aligned}$$

2. **(10 marks)** What is the expected number of cars the traffic officer has to watch to achieve her quota? What distributional are you using and why? *We are assunming that drivers behaviour is independent. Let* T denote the number of traffic tickets

$$\begin{aligned} P(T) &= .2 \\ E[T] &= n \times p = n \times .2 = 4 \Longrightarrow n = 20 \end{aligned}$$

3. (**10 marks**) Using the number of vehicles to reach quota from (2) what is the probability that between 4 and 8 inclusive fail to make the legal stop.

$$\begin{aligned} P(4 &\leq T \leq 8|n = 20, p = .2) = P(T \leq 8) - P(T \leq 3) \\ &= .99 - .206 = .784 \end{aligned}$$

4. **(10 marks)** Approximate the distribution you used in Question 3 with a normal distribution without any corrections. What are the problems associated with this approximation? *Using the normal approximation without any corrections will lead to too small an estimate. We are not taking into account that the binomila distribution is a discrete distribution and that the $P(T = 4)$ and $P(T = 8)$* for discrete distribution is not 0.

$$\begin{aligned} T &\sim N(np, np(1-p)) \\ P(4 &\leq T \leq 8) \approx P\left(\frac{4-4}{\sqrt{20 \times .2 \times .8}} \leq \frac{T - np}{\sqrt{np(1-p)}} \leq \frac{8-4}{\sqrt{20 \times .2 \times .8}}\right) \\ &= P(0 \leq Z \leq 2.24) = F_Z(2.24) - .5 = .9875 - .5 = .4875 \end{aligned}$$

5. **(10 marks)** Improve upon the probability calculation in Question 4 with some corrections. Why does this help? *When we add and subtract .5 we widen the interval and therby increase the probability. This helps correct for the 0 probability assigned at the end points for continuous (normal here) distributions*

$$\begin{aligned} T &\sim N(np, np(1-p)) \\ P(4 - 0.5 &\leq T \leq 8 + 0.5) \approx P\left(\frac{3.5-4}{\sqrt{20 \times .2 \times .8}} \leq \frac{T - np}{\sqrt{np(1-p)}} \leq \frac{8.5-4}{\sqrt{20 \times .2 \times .8}}\right) \\ &= P(-.280 \leq Z \leq 2.516) = (F_Z(.280) - .5) + ((F_Z(2.516) - .5 \\ &= .9941 + .6103 - 1 = .4875 = .6044 \end{aligned}$$

:

Part II (50 marks) There are 5 questions in total for this part. The section tests your understanding of estimation and sampling theory.

Two normal, identically and independently distributed populations of tax returns (measured in dollars) from Ontario (O) and Quebec (Q) are

$$O \sim N_i(15,000, 2000) \text{ and } Q_j \sim N(13,000, 3000)$$

6. **(10 marks)** A research takes a random (independent) sample of 20 from Ontario and 15 from Quebec and wishes to estimate the difference in tax returns but for some reason uses only the 5th observation for the estimation

$$D = O_5 - Q_5$$

What is the distribution, expected value and variance of D? *Since D is a linear combination of independent normally distributed variables it is also normally distributed.*

$$\begin{aligned} D &\sim N(2,000, 5000) \\ E[D] &= E[O_5] - E[Q_5] = 2,000 \\ V[D] &= V[O_5] + V[Q_5] = 5000 \end{aligned}$$

7. (**10 marks**) Having taken Economics 250 you know that you can do better. What is the best estimator of the **population difference?** What is its mean and variance. *we know that the sample mean is the minimum variance unbiased estimator of the population mean. therefore the difference of the two sample means is the minimum variance unbiased estimator of the population difference. We can see clearly below that the variance of this estimator is much less than for D* above

$$\begin{aligned} (\bar{O}_{20} - \bar{Q}_{15}) &\sim N(2,000, 300) \\ V\left[(\bar{O}_{20} - \bar{Q}_{15})\right] &= \frac{2000}{20} + \frac{3000}{15} = 300 \end{aligned}$$

8. (**10 marks**) Calculate the probability of the difference in the researchers estimator of the difference in tax returns being between 1900 and 1995

$$\begin{aligned} P(1900 &< D < 1995) = P\left(\frac{1900 - 2000}{\sqrt{5000}} < \frac{D - \mu_D}{\sqrt{V[D]}} < \frac{1995 - 2000}{\sqrt{5000}}\right) \\ &= P(-1.414 < Z < -.07) = F_Z(1.414) - F(.07) \\ &= .9207 - .5279 = .3928 \end{aligned}$$

9. (**10 marks**) Calculate the probability of the difference in your estimator in the difference in tax returns being between 1900 and 1995.

$$\begin{aligned} P(1900 &< (\bar{O}_{20} - \bar{Q}_{15}) < 1995) = P\left(\frac{1900 - 2000}{\sqrt{300}} < \frac{(\bar{O}_{20} - \bar{Q}_{15}) - \mu_{(\bar{O}_{20} - \bar{Q}_{15})}}{\sqrt{V[(\bar{O}_{20} - \bar{Q}_{15})]}} < \frac{1995 - 2000}{\sqrt{300}}\right) \\ &= P(-5.7735 < Z < -.289) = 1 - F_Z(.289) \\ &= 1 - .6141 = .3859 \end{aligned}$$

10. (**10 marks**) Show in a diagram the results for 8 and 9 and explain. *The difference in the sample means is much more tightly distributed around the population difference (2000). Since this question is about an interval that is away from the mean it is less likely to occur.*

Formula Sheet

Statistics Formulas

Notation

- All summations are for $i = 1, \ldots, n$ unless otherwise stated.
- $\sim$ means 'distributed as'

Population Mean

$$\mu \; = \; \frac{1}{N}\sum_{i=1}^{N} x_i$$

Sample Mean

$$\bar{X} \; = \; \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Population Variance

$$\begin{aligned} \sigma^2 \quad &= \quad \frac{1}{N}\sum_{i=1}^{N}[x_i \; - \; \mu]^2 \\ &= \quad \frac{1}{N}\sum_{i=1}^{N} x_i{}^2 - \mu^2 \end{aligned}$$

Sample Variance

$$s^2 \quad = \tfrac{1}{n-1}\textstyle\sum_{i=1}^{n}[x_i \; - \; \bar{X}]^2$$

Alternatively

$$s^2 = \frac{1}{n-1}[\sum_{i=1}^{n} x_i^2 \; - \; n\bar{X}^2]$$

Grouped Data (with k classes)

$$\begin{aligned}\bar{X} &= \frac{1}{n}\sum_{j=1}^{k}\nu_j f_j \quad \text{where } \nu_j \text{ is the class mark for class } j \\ s^2 &= \frac{1}{n-1}\sum_{j=1}^{k} f_j(\nu_j - \bar{X})^2\end{aligned}$$

Probability Theory

$$\begin{aligned}P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad \text{(additive law)} \\ P(A \cap B) &= P(B)P(A \mid B) \quad \text{(multiplicative law)}\end{aligned}$$

If the E_i are mutually exclusive and exhaustive events for $i = 1, \ldots, n$, then

$$\begin{aligned}P(A) &= \sum_{i}^{n} P(A \cap E_i) = \sum_{i}^{n} P(E_i)P(A \mid E_i) \\ P(E_i &\mid A) = \frac{P(E_i)P(A \mid E_i)}{P(A)} \quad \text{(Bayes' Theorem)}\end{aligned}$$

Counting Formulae

$$P_R^N = \frac{N!}{(N-R)!}$$

$$C_R^N = \binom{N}{R} = \frac{N!}{(N-R)!R!}$$

Random Variables

Let X be a discrete random variable, then:

$$\begin{aligned}
\mu_x &= E[X] = \sum_x xP(X=x) \\
\sigma_x^2 &= V[X] = E[(x-\mu_x)^2] = \sum_x (x-\mu_x)^2 P(X=x) \\
\sigma_x^2 &= E[X^2] - (E[X])^2
\end{aligned}$$

The **covariance** of X and Y is

$$\begin{aligned}
Cov[X,Y] &= \sigma_{xy} = E\left[(X-\mu_x)(Y-\mu_y)\right] \\
&= E[XY] - E[X] \times E[Y]
\end{aligned}$$

The **correlation coefficient** of X and Y

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

If X and Y are independent random variables and a, b, and c are constants, then:

$$\begin{aligned}
E[a+bX+cY] &= a + b\mu_x + c\mu_y \\
V[a+bX+cY] &= b^2\sigma_x^2 + c^2\sigma_y^2
\end{aligned}$$

If X and Y are correlated then

$$V[a+bX+cY] = b^2\sigma_x^2 + c^2\sigma_y^2 + 2bc\sigma_{xy}$$

Coefficient of Variation (CV)

$$
\begin{aligned}
CV &= \frac{\sigma}{\mu} \times 100\% \text{ for population} \\
CV &= \frac{s}{\bar{X}} \times 100\% \text{ for sample}
\end{aligned}
$$

Univariate Probability Distributions

Binomial Distribution: For $x = 0, 1, 2, ..., n$ and :

$$Pr[X = x] = \binom{n}{x} \pi p^x (1-p)^{n-x}$$

$$E[X] = np$$

$$V[X] = np(1-p)$$

Uniform Distribution: For $a < x < b$:

$$f(x) = \frac{1}{b-a}$$

$$E[X] = \frac{a+b}{2}$$

$$V[X] = \frac{(b-a)^2}{12}$$

Normal Distribution:For $-\infty < x < \infty$:

$$f(x) = (2\pi\sigma^2)^{-1/2} exp[\frac{-(x-\mu)^2}{2\sigma^2}]$$

$$E[X] = \mu$$

$$V[X] = \sigma^2$$

$$
\begin{aligned}
X &\sim N(\mu_X, \sigma_X^2) \\
Z &= \frac{X - \mu_X}{\sigma_X} \sim N(0,1)
\end{aligned}
$$